Teamxyz-Report

by Harshit Jain

Submission date: 23-Nov-2022 02:58PM (UTC+0530)
Submission ID: 1961946597
File name: Teamxyz-Report.pdf (95.34K)
Word count: 2313
Character count: 12384

Autism Spectrum Disorder in Adults

1st Karuna K Chandra
Department of CSE)
PES University
Bangalore, India
karunakkc@gmail.com

2nd Lenver Pinto
Department of CSE)
PES University
Bangalore, India
lenverpinto@gmail.com

3rd Harshit Jain
Department of CSE)
PES University
Bangalore, India
hjain1523@gmail.com

***Abstract*—Early diagnosis can dramatically save healthcare expenses associated with autistic spectrum disorder (ASD), a neuro developmental illness. Datasets on behavioral features are required due to the worldwide rapid increase in ASD cases. It is difficult to conduct in-depth study to improve the efficacy, prediction accuracy and sensitivity of the ASD-screening technique due to the scarcity of such datasets. There are currently very few clinical or screening-related autism datasets available, and the majority of them are genetic in nature.**

***Index Terms*—ASD, clinical, screening, comparative study, machine learning, neural networks, data analytic**

I. Introduction

Several Variations in a brain are found to be the root cause of the developmental disorder known as Autism Spectrum Disorder(ASD). ASD patients may struggle with repetitive behaviours or hobbies, and also with social interaction and communication. People with ASD might also move around, grasp, or focus in different ways. It must be kept in mind that people without ASD can also experience some of the above symptoms. Nevertheless, these traits can make life extremely difficult for persons with ASD.

ASD diagnosis procedures are expensive and have extensive waiting periods. The growing prevalence of ASD cases worldwide and the economic cost of autism screening point to the urgent need for the development of simple-to-use screening techniques. In order to assist healthcare practitioners and inform patients about the need for formal clinical diagnosis, a quick and easy-to-use ASD screening is needed. The most common dataset used in research has 20 features and 704 records. The ten behavioral traits (AQ-10-Adult) and ten person characteristics that behaviorists have found to be most useful in seperating ASD cases from controls are noted in this dataset.

The dataset has been used in several research practices which involve screening for autism. Using this dataset, we have developed 4 machine learning and neural network models to assess the effectiveness of each of the model. Based on the accuracy and other parameters, the most relevant model can be used.

II. Literature Survey

In study paper [1], the following data were collected: (1) the sampling method; (2) descriptive variables (such as age, gender, etc); (3) the methods used to diagnose ASD; (4) the percentage of sample members who had an identification number; (5) the methods used for measuring anxiety and depression; (6) if overshadowing/symptom overlap of diagnosis was taken into consideration in the study; and (7) the lifetime and current estimates of anxiety. A prevalence of 42% was found after further analysis of the eight studies that were classified as assessing lifetime expected prevalence. The findings of twelve studies on the topic revealed a lifetime expected prevalence of 20% and a present prevalence of 29% for social anxiety. OCD rates were evaluated across 15 investigations, with lifetime expected prevalence estimates of 22% and current prevalence estimates of 24%. The present frequency of GAD was confirmed by seven investigations.

In a second study paper [2] 32 adults (all male) with ASD were chosen from mailing lists of high-function people with ASD and from local mental health groups. Six ASD sufferers were chosen from a nearby forensic clinic. No participant had an IQ score below 70, and their clinicians regarded them as high-functioning.A clinical psychologist or psychiatrist made an ASD diagnosis for each participant in accordance with DSM-IV-TR criteria after reviewing their developmental history, present-day functioning, and observation. The results showed that the ADOS classification level had good or high interrater agreement and a Cohen's adjusted weighted kappa of 81.7 percent. . With kappa 0.73 after combining the ADOS-classifications for AD and ASD, agreement reached 89.2%. High inter-rater agreement on SOC and COMSOC, as well as good agreement on COM, are revealed by intraclass correlations. With a mean weighted kappa of 0.66, the items had an average agreement of 81.7%. Internal consistency for the original approach is good for SOC (Cronbach's. = 0.84), but somewhat low for COM (= 0.52). This shows that in our population, the things from that domain do not correspond well.

Reference [3] discusses many approaches previously used to categorize ASDs as well as how computer vision has evolved to be useful in ASD diagnosis and autism research more generally. They provide a comparison of deep learning methods with more conventional machine learning techniques, which helped us choose the best course of action.

Another study by other authors for reference [4] offers a thorough understanding of employing SVM classifiers for early and accurate classification of ASD.The authors use adaboost, linear discriminant analysis, and logistic regression in addition to multi-variate correlation and feature selection as machine learning techniques to obtain up to 100% accuracy and other critical parameters. The study gives us some useful information that will help us choose our models intelligently and keep computational power in mind while we do so.

Study [5] summarises the various fMRI research findings using data from the Autism Brain Imaging and Data Exchange and talks about how machine learning approaches and deep learning techniques have evolved over the past three years for the classification and recognition of ASDs (ABIDE). It is discussed how well the two categorization and recognition approaches for ASD perform, which can be helpful for making a premature diagnosis of ASD patients. It also provides the current issues and development trends. These metrics consist of precision, specificity and sensitivity.

The proposed methods in study [6] are assessed on three unique publicly accessible ASD datasets that are not therapeutically relevant. The first dataset relates to the screening for ASD in children and contains 292 cases and 21 attributes.The second dataset for ASD screening consists of adult adults and comprises of 704 cases and 21 attributes.Dataset 3 focuses on ASD screening in adolescents and contains 104 cases and 21 features. The results of the ASD Screening for Adults, Children, and Adolescents strongly imply that CNN-based prediction models outperform them all on these datasets, with accuracies of 99.53%, 98.30% and 96.88%, respectively. After managing missing values and utilising several machine learning approaches, this is the result.

III. Methodology

The aim of this paper is to determine the best model to determine whether an individual will have Autism, based on multiple parameters such as age, ethnicity, nationality, gender, family history of autism, and the answers to the questions, all of which are recorded in the dataset.

A. Dataset

The dataset has 20 features or attributes and 704 observations or records.The attributes include: A1-score, A2-score, A3-score, A4-score, A5-score, A6-score, A7-score, A8-score, A9-score, A10-score, age, gender, ethnicity, jaundice, autism history, if used the app before, country of residence, result of test, age description, relation of the person filling the form, class/ASD.

B. Exploratory Data Analysis

The dataset had few missing values, outliers and junk values. For the outlier, the record was dropped whereas the missing values were replaced by the mean of the attribute they belonged to (age).
The ethnicity attributes had many junk values and data inconsistencies, all of which were replaced by a common value 'Others' which already existed for few records in the dataset. Dropping the records would have decreased the dataset size further which would lead to major issues during the model training. Similarly, for ethnicity the junk values were replaced by the mode.
The dataset was visualized using pie charts for few categories such as gender, classes labeled YES for ASD and NO for ASD, and percentage gender for class labelled as YES.
Since the dataset had several categorical attributes, it was important to ensure that they were suitably encoded in order to use those attributes for computation. Therefore, all the categorical variables were one hot encoded. For those with only 2 possible values, after encoding, one of the columns were dropped to decrease the number of features.
On plotting the correlation matrix, we could determine the most important features which have maximum impact on the response feature.

TABLE I
Features Selected

Feature	*Correlation*
A1 score	0.3
A2 score	0.31
A3 score	0.44
A4 score	0.47
A5 score	0.54
A6 score	0.59
A7 score	0.35
A8 score	0.24
A9 score	0.64
A10 score	0.39
result	0.82

C. Baseline Models

In the baseline modelling, the entire dataset with all its 20 features were used for training after cleaning it. The models used for training the dataset were- Artificial Neural Network, Support Vector Machine, Gradient Descent, and Naive Bayes. Using the scikit-learn python library, the model training was carried out. From the dataset, 80% is used for training and 20% for testing. There is no validation dataset.

TABLE II
Baseline Model Accuracy

Model	*Accuracy*
Artificial Neural Network	74.46%
Support Vector Machine	96.45%
Gradient Descent	96.45%
Naive Bayes	37.59%

It is clear from the table 2 and the graph in Fig 1 that the models which performed the best were Support Vector Machine and Gradient Descent with an accuracy of 96.45%. Naive Bayes performed the most poorly with an accuracy of 37.59%.

D. Feature Selected Models

On training the model through the entire dataset, the above accuracy were obtained.To further improve the model perfor-

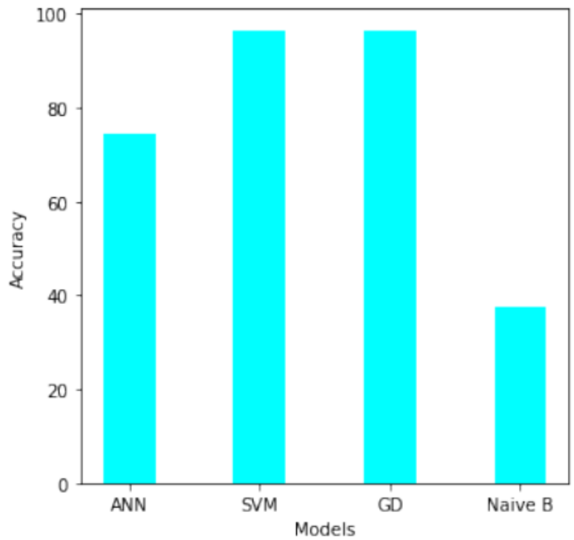

Fig. 1. Baseline Model Accuracy

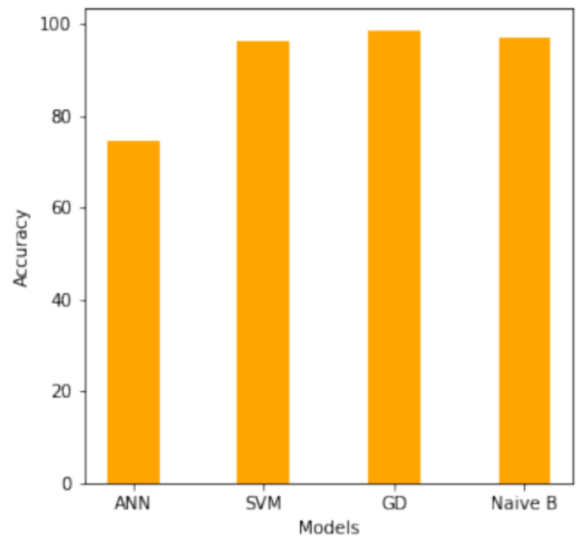

Fig. 2. Baseline Model Accuracy

mance, only few of the features out of the 20 features are selected for training such that only the attributes with the most correlation to the response variable are selected. Once again, from the dataset, 80% is used for training and 20% for testing. There is no validation dataset. The features selected are the above mentioned features in Table 1 which have the most correlation to the response variable.

TABLE III
BASELINE MODEL ACCURACY

Model	***Accuracy***
Artificial Neural Network	74.46%
Support Vector Machine	96.45%
Gradient Descent	98.58%
Naive Bayes	97.16%

From the above graph in Fig 2 and the table 3, we can see that feature selection has had different effects on different models. On Artificial Neural Networks and Support Vector Machine, feature selection did not have much of an effect and the accuracy remained the same. In case of Gradient Descent, the accuracy improved by 2.13%. The most surprising result was from Naive Bayes where the accuracy shot up to 97.16% from a mediocre accuracy of 37.59% in the baseline model, making it the second most suitable model after gradient descent.

E. Hyper Parameter Tuning

Since 3 out of the 4 models showed an accuracy of above 95%, hyper parameter tuning was not found necessary to perform as there isn't much need to improve the accuracy further.

F. Comparison of Baseline and Feature Selected Models

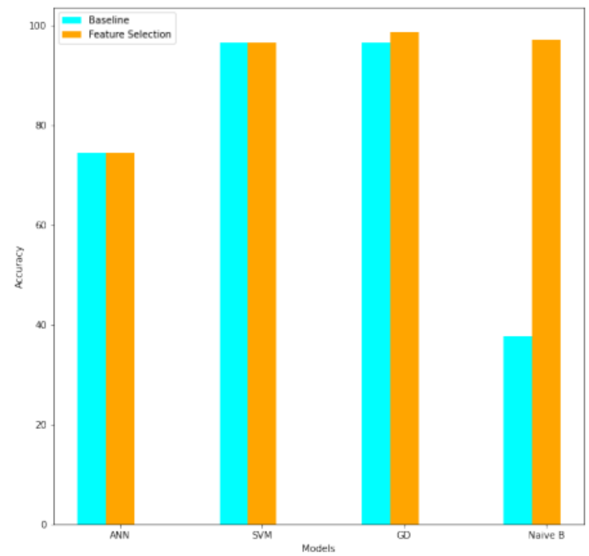

Fig. 3. Baseline Model Accuracy

IV. RESULTS

Training and testing of the models were carried out in 2 stages, baseline modelling where all the features were considered and the feature selection modelling where select features were considered. When it came to baseline model, Support Vector Machine and Gradient Descent were best

with a very high accuracy of 96.45%. But when feature selection was performed, SVM performance did not improve by much, where as Naive Bayes outperformed SVM after an improved accuracy of almost 60%. Surprisingly, ANN model had average performance for both the baseline and the feature selection modelling.
The results of these models and their accuracy may not be completely accurate due to the small dataset size which leads to several issues such as over-fitting. The dataset has only 704 records which is further split into training and testing dataset, leaving only around 560 data records for training which is not sufficient.

V. CONCLUSION

From the above modelling we could conclude that gradient descent is the most suitable model to predict whether an individual has autism spectrum disorder or not. Usage of machine learning and artificial intelligence should always be followed with caution since the misdiagnosis of an individual can have a large negative influence on the individual and their life. Therefore in order to build a more accurate and reliable model it is important to have a large, usable dataset which the model can be trained on resulting in more accuracy. Smaller datasets are only useful to an extent, beyond which there is always the chance of a model being over-fitted for that dataset.

REFERENCES

[1] Hollocks MJ, Lerh JW, Magiati I, Meiser-Stedman R, Brugha TS. Anxiety and depression in adults with autism spectrum disorder: a systematic review and meta-analysis. Psychol Med. 2019Mar;49(4):559-572. doi:10.1017/S0033291718002283. Epub 2018 Sep 4. PMID: 30178724.

[2] Bastiaansen JA, Meffert H, Hein S, Huizinga P, Ketelaars C, Pijnenborg M, Bartels A, Minderaa R, Keysers C, de Bildt A. Diagnosing autism spectrum disorders in adults: the use of Autism Diagnostic Observation Schedule (ADOS) module 4. J Autism Dev Disord. 2011 Sep;41(9):1256-66. doi: 10.1007/s10803-010-1157-x. PMID: 21153873; PMCID: PMC3156304.

[3] Srishti Rau, Mary F. Skapek, Kaitlyn Tiplady, Sydney Seese, Alison Burns, A. Chelsea Armour, Lauren Kenworthy, Identifying comorbid ADHD in autism: Attending to the inattentive presentation, Research in Autism Spectrum Disorders,Volume 69, 2020,101468, ISSN 1750-9467

[4] S. Karim, N. Akter, M. J. A. Patwary and M. R. Islam, "A Review on Predicting Autism Spectrum Disorder(ASD) meltdown using Machine Learning Algorithms," 2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), 2021, pp. 1-6, doi: 10.1109/ICEEICT53905.2021.9667827.

[5] Weibin Feng, Guangyuan Liu, Kelong Zeng, Minchen Zeng, Ying Liu, A review of methods for classification and recognition of ASD using fMRI data, Journal of Neuroscience Methods,Volume 368, 2022,109456, ISSN 0165-0270.

[6] Suman Raj, Sarfaraz Masood, Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques, Procedia Computer Science, Volume 167,2020, Pages 994-1004, ISSN 1877-0509

Teamxyz-Report

ORIGINALITY REPORT

11%	4%	8%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Suman Raj, Sarfaraz Masood. "Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques", Procedia Computer Science, 2020 Publication	2%
2	Submitted to Bournemouth University Student Paper	2%
3	G Sanathkumar, K J Nagesh, Guruprasad Hadimani, Laxman, B R Charanraj, Prasad B Honnavalli. "Smart Waste Segregation", 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC), 2021 Publication	1%
4	Submitted to University of Pittsburgh Student Paper	1%
5	www.coursehero.com Internet Source	1%
6	Matthew J Hollocks, Jian Wei Lerh, Iliana Magiati, Richard Meiser-Stedman, Traolach S Brugha. "Anxiety and depression in adults	1%

with autism spectrum disorder: a systematic review and meta-analysis", Psychological Medicine, 2018
Publication

---

7 www.researchsquare.com <1%
Internet Source

---

8 Khan Md. Hasib, Farhana Rahman, Rashik Hasnat, Md. Golam Rabiul Alam. "A Machine Learning and Explainable AI Approach for Predicting Secondary School Student Performance", 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), 2022 <1%
Publication

---

9 link.springer.com <1%
Internet Source

---

10 "A hybrid method for false data injection attack detection in smart grid based on variational mode decomposition and OS-ELM", CSEE Journal of Power and Energy Systems, 2020 <1%
Publication

---

11 "Chapter 300005 3-Chloro-5-[3-(dimethylamino)propyl]-10,11-dihydro-5H-dibenz[b,f]azepine Monohydrochloride", Springer Science and Business Media LLC, 2021 <1%
Publication

12	asiair.asia.edu.tw Internet Source	<1 %
13	www.spandidos-publications.com Internet Source	<1 %
14	Haishuai Wang, Lianhua Chi, Hong Yang, Li Li, Ziping Zhao. "A deep learning predictive classifier for autism screening and diagnosis", Elsevier BV, 2021 Publication	<1 %

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		